

PATENT APPLICATION

5 **METHODS, SYSTEMS AND COMPUTER SOFTWARE**

 FOR

 DETECTING PIXEL STUTTER

10 Inventors: **Dan Bartell**
 a Citizen of the United States of America,
 residing at 3321 Brittan Ave., #17
 San Carlos, California 94070
 United States of America

15 **Richard Watts**
 a Citizen of the United States of America,
 residing at 862 W. Hillsdale Blvd.
 San Mateo, California 94403
20 United States of America

 Assignee: **AFFYMETRIX, INC.**
 3380 Central Expressway
 Santa Clara, California 95051
 a Delaware corporation

25 Status: Large Entity

 Legal Department
 Affymetrix, Inc.
 3380 Central Expressway
30 Santa Clara, CA 95051

METHODS, SYSTEMS AND COMPUTER SOFTWARE FOR DETECTING PIXEL STUTTER

COPYRIGHT NOTICE

A portion of the disclosure of this patent document contains material that is
5 subject to copyright protection. The copyright owner has no objection to the
xerographic reproduction by anyone of the patent document or the patent disclosure in
exactly the form it appears in the Patent and Trademark Office patent file or records,
but otherwise reserves all copyright rights whatsoever.

10

APPENDIX

Appendix A is included herewith and forms a part of the disclosure.

BACKGROUND OF THE INVENTION

The present invention relates to the field of imaging. In particular, the
15 present invention provides methods and apparatus for detecting pixel stutter and for
improving scanner hardware and software design and scanner manufacturing quality
control.

U.S. Patent No. 5,424,186 describes a pioneering technique for, among other
things, forming and using high density arrays of molecules such as oligonucleotide,
20 RNA, peptides, polysaccharides, and other materials. This patent is hereby
incorporated by reference for all purposes. In this preferred technique, arrays of
oligonucleotides or peptides, for example, are formed on the surface by sequentially

removing a photoremovable group from a surface, coupling a monomer to the exposed region of the surface, and repeating the process. These techniques have been used to form extremely dense arrays of oligonucleotides, peptides, and other materials. Such arrays are useful in, for example, drug development, gene expression monitoring, genotyping, and a variety of other applications.

Genetic mutations or gene expression may be detected by specific nucleic acid hybridization. A solution containing one or more targets to be analyzed (i.e., samples from patients) contacts the probe array. The targets will bind or hybridize with complementary probe sequences. Generally, the targets are labeled with a fluorescent marker, radioactive isotopes, enzymes, or other types of markers. Accordingly, locations at which targets hybridize with complimentary probes can be identified by locating the markers. Based on the locations where hybridization occur, information regarding the target sequences can be extracted. The existence of a mutation may be determined by comparing the target sequence with the wild type.

A number of scanning systems, such as the GeneArray® scanner (Affymetrix and Agilent, California) and 428™ scanners (Affymetrix, California), are available for obtaining hybridization images. However, there is still need in the art for methods, software and systems that evaluate the performance of scanning systems.

SUMMARY OF THE INVENTION

In one aspect of the invention, computer implemented methods are provided to detect pixel stutter. "Pixel stutter" is two or more adjacent pixels with the same

intensity value because of scanning or other artifact(s).

In an exemplary embodiment, a numeric count of stuttered pixels in both the x (row) and y (column) directions and a graphic display of the stuttered pixels overlaid on the image data are provided by the invention. One of skill in the art would

5 appreciate that the designation of rows vs. columns and x vs. y direction are for the convenience of describing embodiments of the invention. In general, the designation of row or x direction indicates that it corresponds to the movement of a scanning device. The y-direction stutter count provides the estimate of pixel stutter from pure statistical data variations. The pixels are measured by the scanner progressing in the
10 x-direction and are possibly correlated to the previous pixel. The pixels in the y-direction are not correlated to the previous pixel. In some embodiments, it is not necessary to provide an estimate of statistical pixel stutter using statistical calculations because the y-direction stutter is a physical measurement of this variation.

15 In preferred embodiments, the methods include obtaining an image comprising a plurality of pixels using a scanner, obtaining a measurement of pixel stutter in the image, obtaining a statistical distribution of pixel stutter, and determining whether said measurement is above what is expected from said statistical distribution. As used herein, an image is a collection of data representing pixel values. An image may be in
20 any suitable format, such as TIFF and BMP.

In some embodiments, the measurement of pixel stutter is time correlated. If the scanner obtains an image row by row, measurement of pixel stutter is row

stuttered pixel count (or horizontal pixel count). In some embodiments, the statistical distribution of pixel stutter is a measurement of non-time correlated pixel stutter in an image and may be measured by column stuttered pixel count (or vertical stuttered pixel count).

- 5 The row stuttered pixel count and column stuttered pixel count can be compared to determine whether the image indicates pixel stutter of the scanner used to obtain it. This may involve a simple calculation of a ratio of horizontal count over vertical count. In some other embodiments, the comparison may be determined by examining statistical difference between said row and said column stuttered pixel
- 10 counts.

In some preferred embodiments, graphic display of stuttered pixels overlaid on the image is provided.

- In another aspect of the invention, computer software products for determining pixel stutter of a scanner are provided. The software products include computer code
- 15 for obtaining an image comprising a plurality of pixels using a scanner; computer program code for obtaining a measurement of pixel stutter in the image; computer program code for obtaining a statistical distribution of pixel stutter; computer program code for determining whether the measurement is above what is expected from the statistical distribution; and a computer readable media for storing the codes.

- 20 In some embodiments, the measurement of pixel stutter is time correlated. The measurement of pixel stutter may be a row stuttered pixel count (or horizontal stuttered pixel count). The statistical distribution of pixel stutter can be a

measurement of non-time correlated pixel stutter such as column stuttered pixel count (or vertical stuttered pixel count).

In some preferred embodiments, the computer program code for determining includes code for comparing row stuttered pixel count and column stuttered pixel
5 count such as the code for comparing includes computer program code for calculating a ratio of horizontal stuttered pixel count over vertical stuttered pixel count or code for comparing comprises computer program code for determining whether there is a statistical difference between said row and said column stuttered pixel counts. The computer software may also include computer program code for displaying stuttered
10 pixels overlaid on the image.

In yet another aspect, systems for detecting pixel stutter of a scanner are provided. The systems include a processor; and a memory being coupled to the processor, the memory storing a plurality of machine instructions that cause the processor to perform a plurality of logical steps when implemented by the processor,
15 the logical steps including: obtaining an image comprising a plurality of pixels using the scanner; obtaining a measurement of pixel stutter in the image; obtaining a statistical distribution of pixel stutter; and determining whether the measurement is above what is expected from the statistical distribution. In some embodiments, the measurement of pixel stutter is time correlated. The measurement of pixel stutter may
20 be a row stuttered pixel count (or horizontal stuttered pixel count). The statistical distribution of pixel stutter can be a measurement of non-time correlated pixel stutter such as column stuttered pixel count (or vertical stuttered pixel count).

In some preferred embodiments, the logic step of determining include comparing row stuttered pixel count and column stuttered pixel count, such as calculating a ratio of said two counts. In some other preferred embodiments, comparison is performed by determining whether there is a statistical difference
5 between said row and said column stuttered pixel counts.

The logic steps may also include displaying stuttered pixels overlaid on the image.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form a part of this
10 specification, illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention:

Figure 1 illustrates an example of a computer system that may be utilized to execute the software of an embodiment of the invention.

Figure 2 illustrates a system block diagram of the computer system of Fig. 1.

15 Figure 3 shows a computer implemented process for counting stuttered pixels.

Figure 4 shows a computer implemented process for detecting pixel stutter of a scanner.

Figure 5 shows stuttered pixels.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

20 Reference will now be made in detail to the preferred embodiments of the invention. While the invention will be described in conjunction with the preferred embodiments, it will be understood that they are not intended to limit the invention to

these embodiments. On the contrary, the invention is intended to cover alternatives, modifications and equivalents, which may be included within the spirit and scope of the invention. All patents and articles are hereby incorporated by reference in their entireties.

5 As will be appreciated by one of skill in the art, the present invention may be embodied as a method, data processing system or program products. Accordingly, the present invention may take the form of data analysis systems, methods, analysis software and etc. Software written according to the present invention is to be stored in some form of computer readable medium, such as memory, hard-drive, DVD ROM
10 or CD ROM, or transmitted over a network, and executed by a processor. For a description of basic computer systems and computer networks, see, e.g., Introduction to Computing Systems: From Bits and Gates to C and Beyond by Yale N. Patt, Sanjay J. Patel, 1st edition (January 15, 2000) McGraw Hill Text; ISBN: 0072376902; and Introduction to Client/Server Systems : A Practical Guide for Systems Professionals
15 by Paul E. Renaud, 2nd edition (June 1996), John Wiley & Sons; ISBN: 0471133337.

Computer software products may be written in any of various suitable programming languages, such as C, C++, Fortran and Java (Sun Microsystems). The computer software product may be an independent application with data input and data display modules. Alternatively, the computer software products may be classes
20 that may be instantiated as distributed objects. The computer software products may also be component software such as Java Beans (Sun Microsystems), Enterprise Java Beans (EJB), Microsoft® COM/DCOM, etc.

Fig. 1 illustrates an example of a computer system that may be used to execute the software of an embodiment of the invention. Fig. 1 shows a computer system 1 that includes a display 3, screen 5, cabinet 7, keyboard 9, and mouse 11. Mouse 11 may have one or more buttons for interacting with a graphic user interface. Cabinet 7 preferably houses a CD-ROM or DVD-ROM drive 13, system memory and a hard drive (*see*, Fig. 2) which may be utilized to store and retrieve software programs incorporating computer program code that implements the invention, data for use with the invention and the like. Although a CD 15 is shown as an exemplary computer readable medium, other computer readable storage media including floppy disk, tape, flash memory, system memory, and hard drive may be utilized. Additionally, a data signal embodied in a carrier wave (*e.g.*, in a network including the internet) may be the computer readable storage medium.

Computer systems are known in the art and are commercially available from sources such as IBM, New York, and Dell, Round Rock, Texas, etc.)

Fig. 2 shows a system block diagram of computer system 1 used to execute the software of an embodiment of the invention. As in Fig. 1, computer system 1 includes monitor 3, and keyboard 9, and mouse 11. Computer system 1 further includes subsystems, such as a central processor 51, system memory 53, fixed storage 55 (*e.g.*, hard drive), removable storage 57 (*e.g.*, CD-ROM), display adapter 59, sound card 61, speakers 63, and network interface 65. Other computer systems suitable for use with the invention may include additional or fewer subsystems. For example, another computer system may include more than one processor 51 or a

09734751-121100

cache memory. Computer systems suitable for use with the invention may also be embedded in a measurement instrument or performed using ASIC devices or the like.

The presently preferred method is designed at increasing the accuracy of data acquired from nucleic acid or other arrays. Arrays are typically two-dimensional

5 substrates having biological compounds thereon. They are used for a variety of purposes, including detection of fluorescent compounds attached to the surface of the substrate. Particularly preferred arrays, uses of the arrays and devices to make, read, and interpret the data from arrays are shown in the following U.S. patents which are These arrays, also described as "microarrays" or colloquially "chips" have been
10 generally described in the art, for example, U.S. Pat. Nos. 5,143,854, 5,445,934, 5,744,305, 5,677,195, 6,040,193, 5,424,186 and Fodor et al., Science, 251:767-777 (1991). Each of which is incorporated by reference in its entirety for all purposes.

These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic
15 methods and solid phase synthesis methods. Techniques for the synthesis of these arrays using mechanical synthesis methods, such as ink jet, channel block, flow channel, and spotting methods which are described in, e.g., U.S. Pat. Nos. 5,384,261, and 6,040,193, which are incorporated herein by reference in their entirety for all purposes. Although a planar array surface is preferred, the array may be fabricated on
20 a surface of virtually any shape or even a multiplicity of surfaces. Arrays may be peptides or nucleic acids on beads, gels, polymeric surfaces, fibers such as fiber optics, glass or any other appropriate substrate, see U.S. Patent Nos. 5,744,305,

5,770,358, 5,789,162, 5,708,153, 6,040,193 and 5,800,992, which are hereby incorporated in their entirety for all purposes. Arrays may be packaged in such a manner as to allow for diagnostics or other manipulation of in an all inclusive device, see for example, US Patent Nos. 5,856,174 and 5,922,591, and 5,945,334, which are
5 incorporated herein in their entireties by reference for all purposes. See also U.S. patent application Serial No. 09/545,207 which is incorporated herein in its entirety for all purposes for additional information concerning arrays, their manufacture, and their characteristics. In acquiring data from the preferred embodiments shown above, it is important that the devices and software will acquire the data accurately.

10 The preferred methods, software products and systems of the present invention are particularly suitable for detecting what is termed "pixel stutter" of imaged which represent the hybridization patterns between the presently preferred nucleic acid probe arrays and their targets. In such preferred embodiments, "scanners" image the targets which are bound to an array by detecting fluorescent or
15 other emissions on the target's labels, or by detecting transmitted, reflected, or scattered radiation. These processes are generally and collectively referred to hereafter for convenience simply as involving the detection of "emissions" and can include optical or other types of emission. Various detection schemes are employed depending on the type of emissions and other factors. A typical scheme employs
20 optical or other elements to provide excitation light and to selectively collect the emissions. Also generally included are various light-detector systems employing photodiodes, charge-coupled devices (CCDs), photomultiplier tubes, and similar

09734751.121400

devices to register the collected emissions. For example, a scanning system for use with a fluorescent label is described in U.S. Pat. No. 5,143,854, incorporated by reference above. Other scanners or scanning systems are described in US Patent Nos. 5,578,832; 5,631,734; 5,834,758; 5,981,956 and 6,025,601, and in PCT Application 5 PCT/US99/06097 (published as WO99/47964), each of which is hereby incorporated by reference in its entirety for all purposes. In the preferred embodiment of the invention, an array has many regions where a compound, such as a nucleic acid is located. Each region contains multiple pixels.

The scanning system provides data representing the intensities (and possibly 10 other characteristics, such as color) of the detected emissions, as well as the locations on the substrate where the emissions were detected. The data typically are stored in a memory device in the form of a data file. One type of data file, commonly referred to as an "image file," typically includes intensity and location information corresponding to elemental sub-areas of the scanned substrate and appears as an "image" of the 15 array. The term "elemental" in this context means that the intensities, and/or other characteristics, of the emissions from this area each are represented by a single value.

When displayed as an image for viewing or processing, this information is often represented by elemental picture elements, or pixels. Thus, a pixel has a single value representing the intensity, or other characteristic, of the elemental sub-area of the 20 substrate from which the emissions were scanned. Color can be used to represent a characteristic of a pixel. For example, a scanned elemental sub-area in which high-intensity emissions were detected may be represented by a pixel having high

luminance (hereafter, a "bright" pixel), and low-intensity emissions may be represented by a pixel of low luminance (a "dim" pixel). Alternatively, the chromatic value of a pixel may be made to represent the intensity, color, or other characteristic of the detected emissions. Thus, an area of high-intensity emission may be displayed as a red pixel and an area of low-intensity emission as a blue pixel. As another example, detected emissions of one wavelength at a particular sub-area of the substrate may be represented as a red pixel, and emissions of a second wavelength detected at an adjacent sub-area may be represented by an adjacent blue pixel. Many other display schemes are known.

10 The methods, systems and software of the invention are also useful for detecting pixel stutter caused by scanner artifact. Scanner, as used herein, is intended to include devices that obtain picture elements in a sequential manner. For example, a scanner may obtain pixels in x or y directions of an image by moving a scanning head in a pixel by pixel fashion. However, scanners may also have mechanisms that obtain more than one pixel at a time.

15 Picture elements (pixels) are represented numerically as intensity values or coordinates in a color space. Once images are acquired in computer readable form, they may be "processed" to remove distortions, to reduce noise, to modify color, to accentuate outlines, and so forth. For a general tutorial overview of image processing, see, for example, Russ, John C., The Image Processing Handbook, 2nd Edition (1995), CRC Press, Inc. which is hereby incorporated by reference in its entirety.

In one aspect of the invention, computer implemented methods are provided to detect pixel stutter. "Pixel stutter" is two or more adjacent pixels that may have different intensity values, but they show the same intensity value because of scanning or other artifact(s). Among possible reasons for pixel stutter is uncleared memory
5 buffer. For example, in one implementation of a laser scanner, optical signals are converted into digital signal via an analog digital converter. A memory buffer is used to hold acquired data. The temporary stored data are then read for further processing. In some instances, the buffer is not cleared and the same intensity value for a pixel may be read several times.

10 One of skill in the art would appreciate that pixel stutter can cause poor quality digital image and erroneous biological measurements. Therefore, methods for detecting pixel stutter are useful for detecting design or operational problems of scanners. In one application, the methods of the invention are used to monitor pixel stutter eradication by redesigning scanner electronics or software.

15 In one additional application, the methods of the invention is used to monitor the quality of images. Even a well designed scanner may occasionally produce an image with unaccepted pixel stutter. In preferred embodiments, the software of the invention for detecting pixel stutter may be a part of a scanner control software. Images obtained are analyzed according to the methods of the invention to detect
20 scanning artifact related to pixel stutter. This is particularly useful for obtaining accurate images of biological arrays. In such application, the intensity value of each pixel provides valuable information about certain biological processes. If the

scanning artifact is not detected early, the arrays may be discarded or otherwise not suitable for re-scanning after certain period of time. In some embodiments of the software products of the invention, a user is warned of potential unacceptable pixel stutter and the software accepts command to rescan to avoid the scanning artifact.

5 Pixels with the same intensity may occur because of spatial correlation or statistical distribution. In one aspect of the invention, the methods of the invention include steps to detect whether there is pixel stutter artifact by examining whether stuttered pixels (pixels with the same value) is above what's expected.

Figure 3 shows a computer implemented process for detecting stuttered pixels
10 in one dimension, such as a row or a column of an image. A computer program code is executed to input a first pixel value (31). The first pixel value is used as a reference (iPixValue=first pixel value). Next pixel value is then inputted (32) and compared with the iPixValue (33). If the next pixel value is equal to the iPixValue, a stuttered pixel counter is incremented (34). Otherwise, there is no increment in the stuttered
15 pixel counter. One of skill in the art would appreciate that the stuttered pixel counter registers pixels that have the same values as their neighbors (in this specific embodiment, in one dimension along a row or a column), whether or not they are caused by scanner artifact.

In one aspect of the invention, a numeric count of stuttered pixels in both the x
20 (row) and y (column) directions and a graphic display of the stuttered pixels overlaid on the image data are provided by the invention. One of skill in the art would appreciate that the designation of rows vs. columns and x vs. y direction are for the

convenience of describing embodiments of the invention. In general, the designation of row or x direction indicates that it corresponds to the movement of a scanning device. The y-direction stutter count provides the estimate of pixel stutter from pure statistical data variations. The pixels are measured by the scanner progressing in the
5 x-direction and are possibly correlated to the previous pixel. The pixels in the y-direction are not correlated to the previous pixel.

Figure 4 shows one embodiment of the computer implemented process of the invention. An image file is inputted (41). Stuttered pixels along the rows of pixels in the image (horizontal stutter count) is counted. Thereafter, stuttered pixels in
10 columns (43) is counted (vertical stutter count). A ratio of horizontal stutter count over vertical stutter count is displayed (44).

In preferred embodiments, the methods include obtaining an image comprising a plurality of pixels using a scanner, obtaining a measurement of pixel stutter in the image, obtaining a statistical distribution of pixel stutter, and determining whether
15 said measurement is above what is expected from said statistical distribution.

In some embodiments, the measurement of pixel stutter is time correlated. If the scanner obtains an image row by row, measurement of pixel stutter is row stuttered pixel count (or horizontal pixel count). In some embodiments, the statistical distribution of pixel stutter is a measurement of non-time correlated pixel stutter in an
20 image and may be measured by column stuttered pixel count (or vertical stuttered pixel count).

The row stuttered pixel count and column stuttered pixel count can be

09734751-121100

compared to determine whether the image indicates pixel stutter of the scanner used to obtain it. This may involve a simple calculation of a ratio of horizontal count over vertical count. In some other embodiments, the comparison may be determined by examining statistical difference between said row and said column stuttered pixel counts.

In some preferred embodiments, graphic display of stuttered pixels overlaid on the image is provided.

In another aspect of the invention, computer software products for determining pixel stutter of a scanner are provided. The software products include computer code for obtaining an image comprising a plurality of pixels using a scanner; computer program code for obtaining a measurement of pixel stutter in the image; computer program code for obtaining a statistical distribution of pixel stutter; computer program code for determining whether the measurement is above what is expected from the statistical distribution; and a computer readable media for storing the codes.

In some embodiments, the measurement of pixel stutter is time correlated. The measurement of pixel stutter may be a row stuttered pixel count (or horizontal stuttered pixel count). The statistical distribution of pixel stutter can be a measurement of non-time correlated pixel stutter such as column stuttered pixel count (or vertical stuttered pixel count).

In some preferred embodiments, the computer program code for determining includes code for comparing row stuttered pixel count and column stuttered pixel count such as the code for comparing includes computer program code for calculating

a ratio of horizontal stuttered pixel count over vertical stuttered pixel count or code for comparing comprises computer program code for determining whether there is a statistical difference between said row and said column stuttered pixel counts. The computer software may also include computer program code for displaying stuttered
5 pixels overlaid on the image.

In yet another aspect, systems for detecting pixel stutter of a scanner are provided. The systems include a processor; and a memory being coupled to the processor, the memory storing a plurality of machine instructions that cause the processor to perform a plurality of logical steps when implemented by the processor,
10 the logical steps including: obtaining an image comprising a plurality of pixels using the scanner; obtaining a measurement of pixel stutter in the image; obtaining a statistical distribution of pixel stutter; and determining whether the measurement is above what is expected from the statistical distribution. In some embodiments, the measurement of pixel stutter is time correlated. The measurement of pixel stutter may
15 be a row stuttered pixel count (or horizontal stuttered pixel count). The statistical distribution of pixel stutter can be a measurement of non-time correlated pixel stutter such as column stuttered pixel count (or vertical stuttered pixel count).

In some preferred embodiments, the logic step of determining includes comparing row stuttered pixel count and column stuttered pixel count, such as
20 calculating a ratio of said two counts. In some other preferred embodiments, comparison is performed by determining whether there is a statistical difference between said row and said column stuttered pixel counts.

The logic steps may also include displaying stuttered pixels overlaid on the image.

Appendix A is a computer program named "DAT TOOL". This program was used to analyze image data sets from two scanners, one that exhibited pixel stutter (Scanner H) and one that did not (Scanner G). Table 1 shows the result of stutter counts.

Table 1. Stuttered Pixel Counts

Scanner	Data	Vertical Stutter Count	Horizontal Stutter Count	Ratio Hor/Ver
H	Image H1	5183	5016	.96
G	Image G2	430	18241	42.4
G	Image G3	325	18300	56.3
G	Image G4	323	17491	54.2

The Ratio column in the table indicates, for example, that Images G2, G3, and G4 have about 42, 56, 54 times, respectively, more pixel stutter than would be expected from purely statistical sources. The results indicate that Scanner H does not have detectable stutter artifact; while Scanner G suffers from stutter artifact.

As mentioned above, a preferred embodiment of the invention relates to arrays of biopolymers, called arrays. See U.S. Patent No. 5,445,934 and the other patents referred to above. Arrays have areas or regions that are designed to have one biopolymer, such as a nucleic acid. There are many pixels in each region or area typically. For example, see Figure 5 in which the lower image shows a light rectangular area (a region) formed by dark areas (other regions) with smaller squares therein (pixels).

The above description is illustrative and not restrictive. Many variations of the invention will become apparent to those of skill in the art upon review of this disclosure. Merely by way of example, while the invention is illustrated with particular reference to the evaluation of scanner used for scanning nucleic acid arrays, the methods, software and systems can be used in the evaluation and design of other scanners such as those used for scanning color pictures, etc. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

10 All references including patent and non-patent literature cited herewith are incorporated herein by reference in their entireties for all purposes.

09734751-121100